BITS F464 Machine Learning

Assignment 1

Date : 15th March, 2020

Team Members

Durba Satpathi	2019A7PS0972H
R Adarsh	2019A7PS0230H
Navdeep Singh Narsinghia	2017B5A71675H

Problem 1A Fisher Discriminant

1.1. Model Description and implementation

Fisher Discriminant analysis is a linear binary classification model using dimensionality reduction. We consider the case of 2 classes C_1 , C_2 and given a d dimensional input vector x we project it to one dimension using y= $w^T \cdot x$ where w is a unit vector such that all the projections of points belonging to C_1 are to one side of threshold value v_o and projections of points belonging to C_2 are to the other side.

Given
Training examples - (x_1, t_1), (x_2, t_2),……….., (x_N, t_N)
where

- x_i is a *d* dimensional unit vector for i ∈ {1,2,…...,N}
- t_i is the target attribute (In the the given examples t_i = 0 if x_i ∈ C1 and t_i = 1 if x_i ∈ C2) for i∈{1,2,…...,N}

Procedure
1)We first find the unit vector w such that that it maximizes the difference between the means of projections of the points belonging to C_1 and the means of projections of the points belonging to C_2 and also minimizes the sum of variances of projections of the points belonging to C_1 and variances of projections of the points belonging to C_2 *i.e* maximize the function $\frac{w^T \cdot (m_1 - m_2)}{{s_1}^2 + {s_2}^2}$ subject to $w^T . w = 1$, where

- $m_1 = \frac{1}{N_1} \sum\limits_{n \in C_1} y_n$ *i.e* mean of the projections of points belonging to C_1
- $m_2 = \frac{1}{N_2} \sum\limits_{n \in C_2} y_n$ *i.e* mean of the projections of points belonging to C_2
- N_1 = number of examples belonging to C_1
- N_2 = number of examples belonging to C_2
- $y_n = w^T \cdot x_n$
- ${s_1}^2 = \frac{1}{N_1} \sum\limits_{n \in C_1} (y_n - m_1)^2$ *i.e* within-class variance of the projections of points belonging to C_1
- ${s_2}^2 = \frac{1}{N_2} \sum\limits_{n \in C_2} (y_n - m_2)^2$ *i.e* within-class variance of the projections of points belonging to C_2

On solving the optimization problem, we get $w \propto S_w^{-1} \cdot (M_1 - M_2)$, where

- $M_1 = \frac{1}{N_1} \sum\limits_{n \in C_1} x_n$ *i.e* mean of the points belonging to C_1

- $M_2 = \frac{1}{N_2} \sum_{n \in C_2} x_n$ *i.e* mean of the points belonging to C_2
- $S_w = \frac{1}{N_1} \sum_{n \in C_1} (x_n - M_1) \cdot (x_n - M_1)^T + \frac{1}{N_2} \sum_{n \in C_2} (x_n - M_2) \cdot (x_n - M_2)^T$ *i.e* total within-class covariance matrix

2) We then find the projections of points w *i.e* $c_1 = \{w^T x_n \,|x_n \in C_1\}$, $c_2 = \{w^T x_n \,|x_n \in C_2\}$.

3) We then plot the normal distributions of the points in c_1 ($Y_{c1} \sim$ N($m_1, s_1{}^2$)) and c_2 ($Y_{c2} \sim$ N($m_2, s_2{}^2$))and find their intersection which is the threshold v_o. Thus given a d dimensional vector x' we find the value y'= $w^T \cdot x'$ and classify it as belonging to C_1, if y'> v_o and belonging to C_2 otherwise. Thus, the line y= v_o is the equation of the discriminant in 1-D.

Calculation of point of intersection of the normal distributions:

➔ $\frac{1}{s_1\sqrt{2\pi}} e^{\frac{-1}{2}\frac{(x-m_1)^2}{s_1^2}} = \frac{1}{s_2\sqrt{2\pi}} e^{\frac{-1}{2}\frac{(x-m_2)^2}{s_2^2}}$

➔ -ln(s_1) - $\frac{1}{2}\frac{(x-m_1)^2}{s_1^2}$ = -ln(s_2) - $\frac{1}{2}\frac{(x-m_2)^2}{s_2^2}$

➔ $(\frac{1}{s_1^2} - \frac{1}{s_2^2})x^2 + (\frac{-2m_1}{s_1^2} + \frac{2m_2}{s_2^2})x + \frac{m_1^2}{s_1^2} - \frac{m_2^2}{s_2^2} + 2ln(\frac{s_1}{s_2}) = 0$

The above equation is quadratic in x ($ax^2 + bx + c = 0$) with a= $\frac{1}{s_1^2} - \frac{1}{s_2^2}$, b= $\frac{-2m_1}{s_1^2} + \frac{2m_2}{s_2^2}$ and c= $\frac{m_1^2}{s_1^2} - \frac{m_2^2}{s_2^2} + 2ln(\frac{s_1}{s_2})$ and its roots can be found as x= $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$ one of which gives the value of v_o

4)The equation of the discriminant in the original space thus becomes $w^T \cdot x = v_o$.
If w = [$w_1, w_2, \ldots..w_d$] and x = [$X_1, X_2, \ldots..X_d$], then the equation of the hyperplane becomes $w_1X_1 + w_2X_2 + \ldots\ldots\ldots + w_dX_d - v_o = 0$.

Results

For the given dataset we get

```
w =  [-0.00655686 -0.01823739  0.99981218]
The value of threshold v_o is  -0.3893028020993765
Accuracy  1.0
```

Discriminating line in 1-D

- ➔ y= v_o =-0.3893028020993765
- ➔ where y= $w^T \cdot x$ and w= [-0.00655686 -0.01823739 0.99981218]

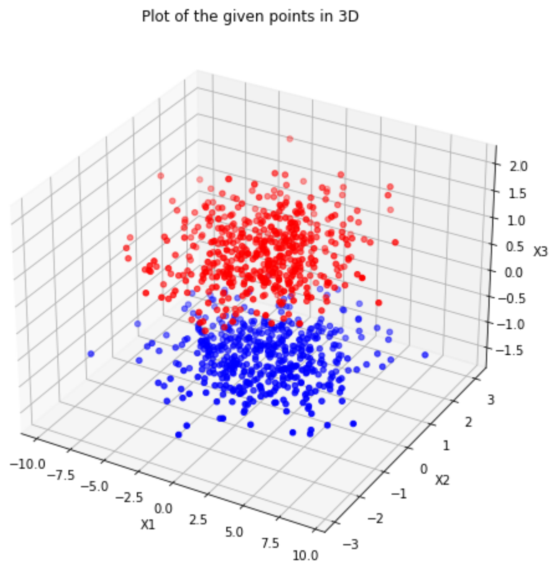
Equation of the discriminating plane in 3-D

➔ w is perpendicular to the discriminating plane;

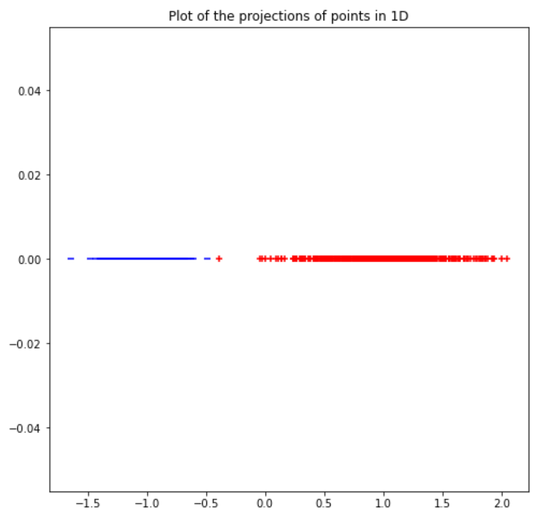
➔ w=`[-0.00655686 -0.01823739  0.99981218]`

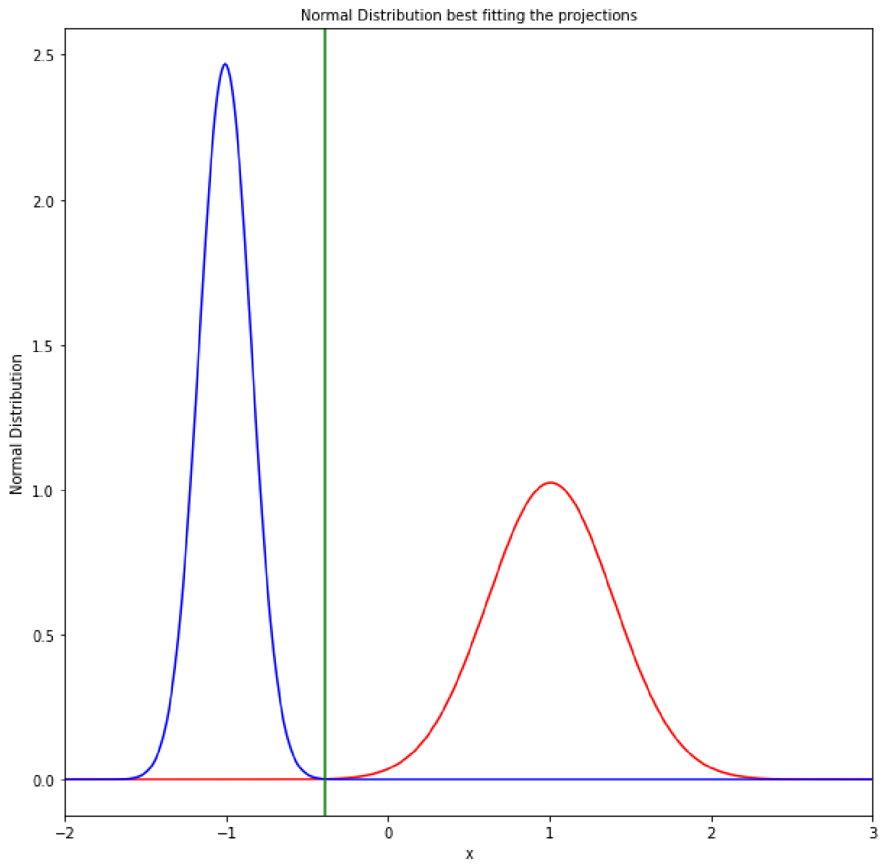
➔ Thus equation of plane becomes -
`-0.00655686(X1)-0.01823739(X2)+0.99981218(X3)=-0.3893028020993765`

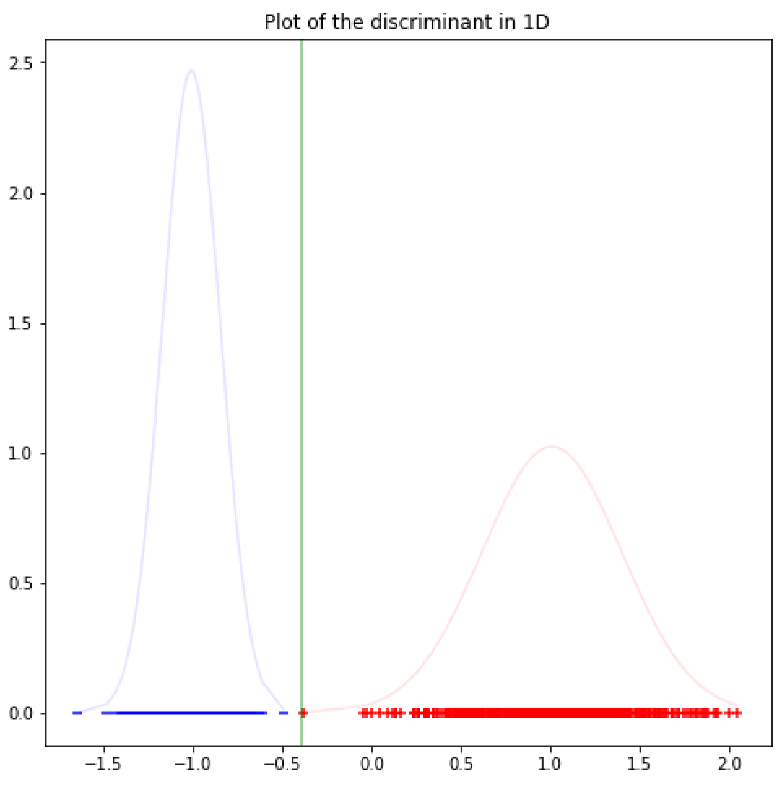
1.2. Plot of the higher dimensional data



1.3. Plots of the reduced clusters and their corresponding normal distribution and unit vector along the discriminant line 1-D

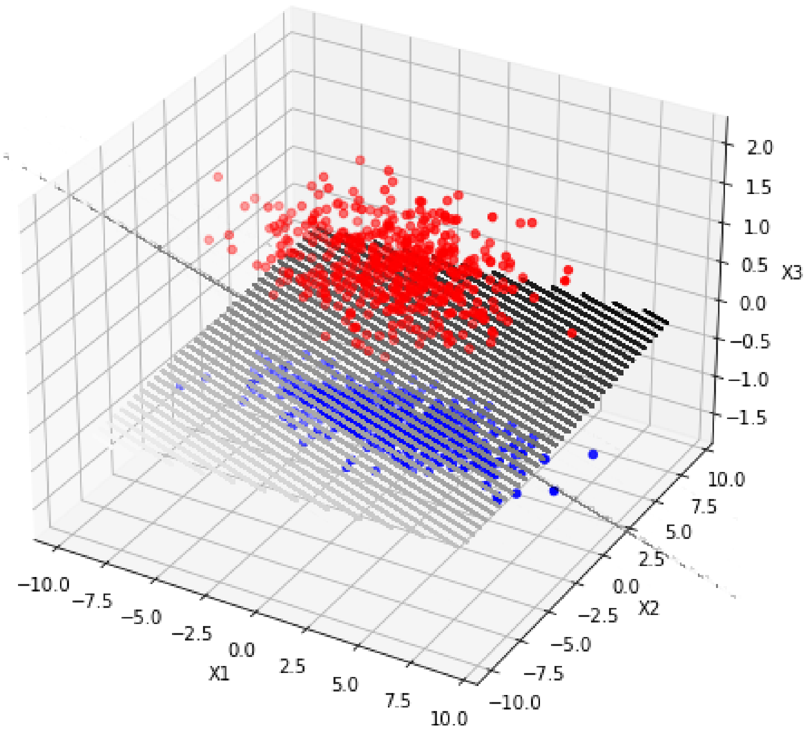



Normal Distribution best fitting the projections
Normal Distribution
x

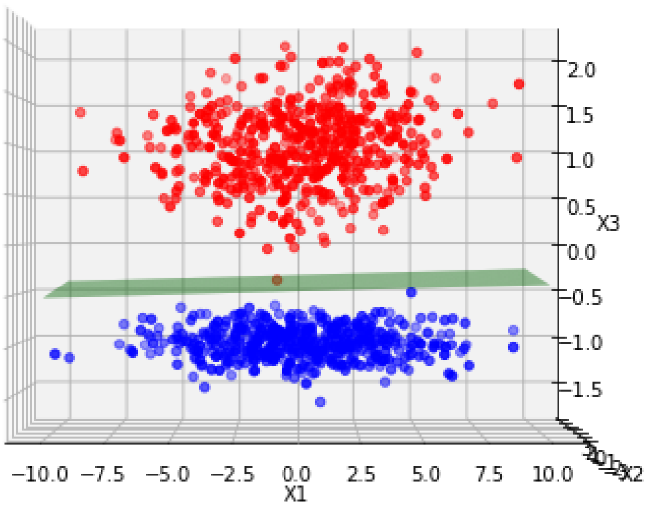

Plot of the discriminant in 1D

1.4.Discriminant in 3-D

Plot of the discriminant in 3D



Plot of the discriminant in 3D at different viewing angle



Projections on w and discriminant in 3D